Pattern Recognition Letters 65 (2015) 170–176

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Cross-pose face recognition based on multiple virtual views and alignment error ☆

Yongbin Gao [a], Hyo Jong Lee [a,b,*]

[a] Division of Computer Science and Engineering, Chonbuk National University, 567 Baekje-Daero, Deokjin-Gu, Jeonju 561756, Republic of Korea
[b] Center for Advanced Image and Information Technology, Chonbuk National University, 567 Baekje-Daero, Deokjin-Gu, Jeonju 561756, Republic of Korea

ARTICLE INFO

Article history:
Received 12 November 2014
Available online 11 August 2015

Keywords:
Face recognition
Virtual views
Alignment error

ABSTRACT

Although studied for decades, effective face recognition remains difficult to accomplish on account of occlusions and pose and illumination variations. Pose variance is a particular challenge in face recognition. Effective local descriptors have been proposed for frontal face recognition. When these descriptors are directly applied to cross-pose face recognition, the performance significantly decreases. To improve the descriptor performance for cross-pose face recognition, we propose a face recognition algorithm based on multiple virtual views and alignment error. First, warps between poses are learned using the Lucas–Kanade algorithm. Based on these warps, multiple virtual profile views are generated from a single frontal face, which enables non-frontal faces to be matched using the scale-invariant feature transform (SIFT) algorithm. Furthermore, warps indicate the correspondence between patches of two faces. A two-phase alignment error is proposed to obtain accurate warps, which contain pose alignment and individual alignment. Correlations between patches are considered to calculate the alignment error of two faces. Finally, a hybrid similarity between two faces is calculated; it combines the number of matched keypoints from SIFT and the alignment error. Experimental results show that our proposed method achieves better recognition accuracy than existing algorithms, even when the pose difference angle was greater than 30°.

1. Introduction

Face recognition has been investigated for several decades. According to the 2009 NIST MBGC report [1], face recognition remains a challenging endeavor on account of variations in poses, illumination, occlusion, and aging. Among these, pose variance is the most difficult to address. Face recognition algorithms can be used to identify criminals from surveillance systems for public security. In addition, they can be applied to automatically annotate digital photos for individuals. Moreover, commercial face recognition software is publicly available, such as Google Picasa and Apple iPhoto [2].

Discriminative facial features are important for both accuracy and speed. Local features, such as local Gabor binary patterns (LGBP) [3] and high-dimensional local binary patterns (LBP) [4], are effective for face recognition. Wright et al. proposed use of sparse representation for face recognition; it can handle illumination, expression variance, and occlusion [5]. The face recognition problem can be divided into two categories: face identification and face verification. Face identification serves to identify a probe face from a set of gallery faces with known identities. Face verification is used to determine whether two images belong to the same subject.

Face verification is a useful branch of the face recognition problem. Recently, a joint Bayesian model trained from a large dataset of labeled faces was successfully used for face verification [6]. Based on this model, a transfer learning method was proposed for combining ample cross-domain source data [7]. In [8], the Facebook AI Research group proposed the DeepFace framework, which uses deep networks for face verification. The system reached a state-of-the-art accuracy of over 97% on the Labeled Faces in the Wild (LFW) database.

Face identification is another challenging endeavor for face recognition. Under controlled settings, such as a frontal face with little illumination variance, the face identification performance has approached human capacity. While pose and illumination variances exist in most applications, face identification accuracy significantly decreases when test faces are non-frontal. Most facial image databases contain only frontal faces, such as driver licenses. The Department of Motor Vehicles collects frontal view images of each driver. Thus, it is necessary to process cross-pose matching to identify a randomly

☆ This paper has been recommended for acceptance by A. Kumar.

* Corresponding author at: Division of Computer Science and Engineering, Chonbuk National University, 567 Baekje-Daero, Deokjin-Gu, Jeonju 561756, Republic of Korea. Tel.: +821026805119.
E-mail address: hlee@chonbuk.ac.kr (H.J. Lee).

http://dx.doi.org/10.1016/j.patrec.2015.07.018

posed face from a frontal view database. Cross-pose identification remains a challenging problem. Moreover, in cases in which only a single frontal face is available, cross-pose matching becomes more difficult. The difficulty in identifying a face with different poses is that the 'between-subject' differences are less than the 'between-pose' differences. There are two solutions for handling this problem: geometry-based methods and pose-invariant-based methods.

The geometry-based method uses an alignment method to build correspondences among different poses. Based on these correspondences, a probe face with different poses can be normalized to a frontal face. This method can be performed in both 2D and 3D cases. Regarding 2D methods, a Markov random field (MRF) is used to find correspondences between a frontal face and profile faces [9,10]. The 2D displacement is captured using MRF by minimizing the energy, which includes the residual of two corresponding nodes and smoothness between neighboring nodes. MRF is effective on some databases; however, it incurs lengthy computation times. Lucas–Kanade is another effective 2D normalization method [11]; it calculates the transformation parameters for each of the correspondences of two images.

The 3D morphable method is an effective normalization method [12] that builds a 3D general model and fits it to a probe 2D face. The fitted shape and texture coefficients are used for face identification. Li et al. [13] synthesized a probe face by estimating 3D displacement fields from a 3D face database. It has been reported that 3D techniques can achieve impressive results on many databases. However, 3D face databases are required for these methods. Furthermore, recovery of a 3D virtual face from a 2D image is difficult because of insufficient information. Moreover, fitting a 3D model to a 2D image is sensitive to factors such as illumination and expression.

The pose-invariant-based method employs pose-invariant features or pose-insensitive classifiers to eliminate the pose influence. Tied-factor analysis has been proposed for representing a non-frontal face by a pose-contingent linear transformation of identifiers [14]. The resultant pose-invariant identity subspace is used for identification. Another subspace, called the discriminant-coupled latent subspace, has been proposed [15]. It is used to find projections of the same subject from different poses that are maximally correlated in the latent subspace. One-shot similarity (OSS) and two-shot similarity (TSS) are pose-insensitive classifiers [16]; they require a third-party dataset with no probe and gallery faces in it. Each subset can be of the same subject with different poses. Similarities between two faces are calculated by models built from these faces and the subsets using linear discriminant analysis (LDA) or support vector machine (SVM). Similarly, cross-pose face recognition likewise requires a third-party dataset [17]. Faces from different poses are linearly represented by the third-party dataset based on a subspace method. The obtained linear coefficients are used for face identification. Recently, neural networks [18] and deep learning [19] have been applied for calculating the pose-invariant features. The networks are trained by converting a non-frontal face to a frontal face; the pose-invariant features are obtained in a specific layer.

In this paper, we propose a novel algorithm based on multiple virtual views and alignment error (MVV–AE) for face recognition under large pose changes with only a single frontal face available for each subject in a gallery. The main contributions of this paper are as follows: (1) scale-invariant feature transform (SIFT)-matching score based on multiple virtual views is proposed to improve the performance of SIFT for the large pose change. A frontal face is transformed into multiple virtual views using learning warps across poses by the Lucas–Kanade algorithm [11]. SIFT is used to calculate the keypoints from these virtual views and to match them to a probe face; (2) a two-phase alignment method is proposed to calculate the alignment error between a probe face and a gallery face. Offline alignment is used to calculate pose differences, while online alignment is used to calculate individual differences. Overlapping and covariance are adopted to capture correlations between patches; (3) a hybrid similarity between probe and gallery faces is obtained by the number of matched keypoints from SIFT and the two-phase alignment error.

The remainder of this paper is organized as follows. In Section 2, we describe the framework of the proposed face recognition algorithm based on MVV–AE. We describe the MVV generation method in Section 3. In Section 4, we introduce the proposed two-phase AE with the correlation method. The combination of these two methods is used for the similarity calculation in Section 5. In Section 6, we apply the above algorithm to the FERET [23] database and present the experimental results. Finally, we present our conclusions in Section 7.

2. Proposed MVV–AE framework

Local features, such as LBP and SIFT, are effective for face recognition with small pose changes. However, the performance significantly decreases with large pose variations. To enable the effectiveness of these local features for large pose changes, we propose a novel framework based on MVV–AE. The challenge of this objective is that gallery faces and probe faces are of different poses and only a single frontal face is stored in the gallery. Fig. 1 shows the framework of the proposed MVV–AE method, whereby warps between poses are learned from numerous face pairs of different poses. Given a gallery face and a probe face, we can calculate their similarity based on two parts: the number of matched keypoints and the alignment error. Since pose estimation is out of the scope of this paper, we assume the pose of a probe face is annotated with the ground truth. The input data for a test consist of a probe face image and its ground-truth pose.

The number of matched keypoints is obtained from the SIFT matching algorithm between the probe face and generated virtual views of a gallery face. For each gallery face, we generate multiple virtual views in advance using the learned warps between poses. SIFT keypoints are detected from these virtual views and matched with the probe face.

The alignment error is calculated by a hybrid alignment method with consideration of the correlations between patches. The differences between two face images are primarily from the pose variance and identity differences. In this study, we consider both of these differences in our hybrid alignment, which contains a pose and individual alignment.

3. Multiple virtual views

Without 3D information, a probe face with a profile angle is difficult to transform into a frontal face because of occlusion, especially in the nose area. However, frontal faces contain intact information with no occlusion. This motivates us to transform a frontal face into multiple profile views instead of normalizing a profile face into a frontal view. In this scheme, transformation of a frontal face into virtual views is performed in advance, while normalization of a profile face into a frontal view should be simultaneously conducted with the face recognition.

To transform a frontal face into multiple virtual views, we must learn warps among multiple poses. As a simple and efficient warp, the affine transformation is used in this study. The human face contains significant 3D depth information; a single affine warp for the entire face is insufficient for capturing transformations between poses. Thus, we divide a face into multiple subregions or patches; a warp is learned for each patch. The Lucas–Kanade algorithm [11] is effective for learning warps between poses, as shown in Fig. 2.

To obtain generic warps, numerous face pairs are used to learn warps between poses. This can be performed by averaging two sets of faces; the averaged face pairs are used to learn warps. However,

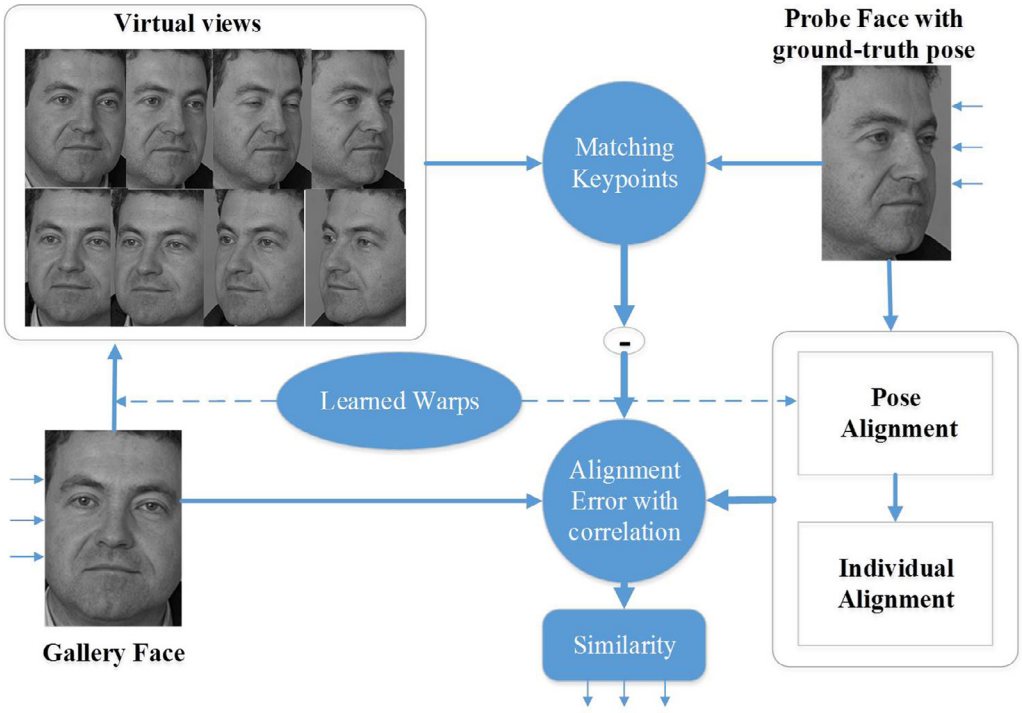

Fig. 1. Framework of proposed MVV–AE algorithm. Based on the learned warps, virtual views are generated from a gallery face. The number of matched keypoints is obtained by comparing virtual views and the probe face. The alignment error is combined with the number of matched keypoints for obtaining the similarity between the gallery and probe face.

an averaging of faces results in neutralization, which is too generalized to represent a specific individual. The stack flow method [11] provides a better solution by calculating warps that minimize all face pairs from two poses as follows:

$$E_{r(stk)} = \sum_j \sum_x \left(I_{j,r}(W(X,P)) - T_{j,r}(X)\right)^2 \tag{1}$$

where I and T are images from two poses, and j and r are the jth pair of images and the patch index, respectively. $W(X,P)$ is the warp function, which is the affine transformation in this study, and

$$W(X,P) = PX = \begin{pmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{2}$$

The Lucas–Kanade algorithm provides a solution for Eq. (1) by iterating the update of P with ΔP as

$$\Delta P = \mathrm{H}_{(stk)}^{-1} \sum_j \sum_X \left(\nabla I_{j,r} \frac{\partial W}{\partial P} \right)^T \left(T_r(X) - I_r(W(X,P))\right) \tag{3}$$

where $\nabla I_r = (\frac{\partial I_r}{\partial x}, \frac{\partial I_r}{\partial y})$, $\frac{\partial W}{\partial P}$, and H_{img} are the gradient of I_r, the Jacobian of the warp and a pseudo Hessian matrix, respectively.

Let N and M be the number of patches in an image and the number of poses, respectively. Let $\Phi = (P_1, P_2, \ldots, P_N)$ be the warps between two poses. A series of warps, $\Phi_1, \Phi_2, \ldots, \Phi_M$, are learned between the frontal view and profile poses. Based on these warps, a frontal face is transformed into multiple virtual poses. The SIFT keypoints are detected from these virtual views and matched with the probe face. Accordingly, the number of matched keypoints between the probe face and gallery faces is obtained.

The probe face is compared with all generated virtual views because it is not easy to accurately estimate the pose; moreover, images from other poses provide additional facial information. However, considering computation time, an interval is required to maintain sparsity. SIFT is chosen for the features on account of its scale-invariant characteristic and good performance for pose variance

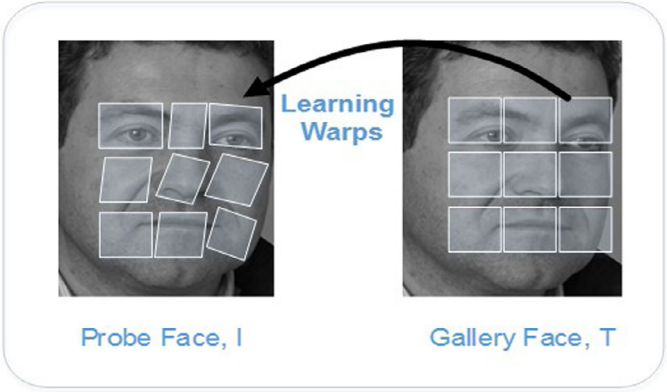

Fig. 2. Learning warps between the probe face and gallery face for each patch. Gallery face is equally divided into several patches, for each patch, a warp is learned using Lucas–Kanade algorithm.

within 25°, according to our experiments. In this regard, poses are divided into four quadrants, we define the *poses* in the same quadrant as the same *orientation*. As for the FERET database, it consists of poses from left to right captured at 60°, 40°, 25°, 15°, −15°, −25°, −40° and −60°. We divided them into two orientations: left and right. Poses at 60°, 40°, 25°, and 15°belong to the same orientation. We generate a virtual view for each pose, and based on our experiments, we found that the virtual views from different orientations of the probe face had a negative impact on the face recognition accuracy (see Section 6.1). Thus, we estimate the orientations (up-left, up-right, down-left, down-right) of the probe face. The virtual views of the same orientation of the probe face are used to calculate the matched keypoints. In this way, the compared virtual views are significantly reduced, and the estimation of orientations is much easier than that of poses.

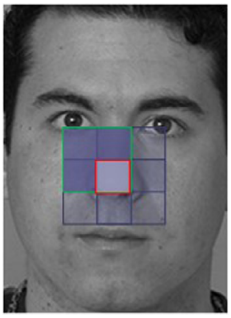

Fig. 3. Illustration of four neighboring patches that overlap: the small center patch is overlapped by four neighboring patches. The warp parameter of the center patch is the average of these four patches.

4. Two-phase alignment error with correlation

Learned warps are the correspondences between poses and are trained from numerous face pairs. When directly applying the learned warp to the probe face, it is not sufficiently accurate on account of individual differences. Once exact warps between faces are available, we can easily calculate the difference between these two faces. The difference of two faces is derived from the pose and individual difference. The learned warps described in Section 3 demonstrate the pose alignment process. To obtain more accurate warps, we propose a two-phase alignment process, which is comprised of the pose alignment and individual alignment. The pose alignment is fulfilled by the learned warps between poses, as outlined in Section 3; it is performed in an offline manner without incurring much computation time. Individual alignment, which is conducted in an online approach, converges quickly because of the pose alignment performed prior to it. Let P_p and P_i be the pose warps and individual warps matrices, respectively. The two-phase alignment error (AE) is used to minimize the following error:

$$E_r = \sum_x \left(I_r(W(X, P_p, P_i)) - T_r(x)\right)^2 \tag{4}$$

where $W(X, P_p, P_i) = P_i P_p X$, and the multiplication of the two-warp matrices, P_p and P_i, corresponds to the pose alignment and individual alignment, respectively. P_p is learned by numerous face pairs of frontal faces and non-frontal faces with ground-truth pose p in advance. For each probe face, an online alignment is performed with each gallery face i using the Lucas–Kanade method by the iteration of update P_i. The online alignment converges quickly due to the pose alignment prior to it. In our experiments, P_i converged after approximately 10–15 iterations, which indicates no significant increase in computational time.

The block effect is incurred during the division of patches. Moreover, the Lucas–Kanade method is performed for each patch separately. The method readily warps adjacent patches to much different areas. Thus, we propose the use of overlapping of patches. Four neighboring patches are overlapped, as shown in Fig. 3. The small patch at the center is overlapped by four neighboring patches; the warp parameter of the center patch is the average of these four patches.

After P_i is obtained, the alignment error can be calculated by Eq. (4) for each patch. Because overlapping is used in our scheme, there is a correlation among patches. Considering these correlations, we multiply the covariance of the alignment error of each patch to calculate the alignment error of the whole face as

$$E^{mah} = \left(E - \bar{E}\right)^T Cov^{-1}\left(E - \bar{E}\right) \tag{5}$$

where $E = (E_1, E_2, \ldots, E_N)$, N is the number of patches, Cov is the covariance of E, $\bar{E}$ is the mean of vector E, Cov and $\bar{E}$ are statistics of alignment errors calculated across the entire gallery faces, E^{mah} is the alignment error of a gallery face and a probe face. Eq. (5) shows that E^{mah} is actually the square of the Mahalanobis distance of E. By using the Mahalanobis distance, the correlations between patches are taken into account in the alignment error.

5. MVV–AE similarity

The number of matched keypoints calculated in Section 3 represents the similarity of two faces, while the alignment error in Section 4 corresponds to the dissimilarity between two faces. We combine these two factors to calculate the similarity index of two faces as follows:

$$S_i = \lambda \frac{M_i}{\max_i (M_i)} - (1 - \lambda) \frac{E_i}{\max_i (E_i)} \tag{6}$$

where M_i and E_i are the number of matched keypoints and the alignment error between the probe face and gallery face, i, respectively. M_i and E_i are normalized to [0,1] by dividing them with the maximum value of M_i and E_i among all the subjects. λ is the weight for these two factors. In general, the proposed MVV–AE face recognition method that is based on matching and the alignment error can be summarized as follows:

Algorithm: MVV–AE similarity

Pre-computation:

1. Learn a series of warp parameters, $\Phi_1, \Phi_2, \ldots, \Phi_M$, between M different poses from a frontal face using M stack images; each stack image is from the same pose.
2. For each frontal face in the gallery, we generate M virtual views based on the learned warp parameters.
3. Compute the SIFT keypoints of these M virtual views and store them in a keypoint database.

Recognition:

1. For each probe face, detect the SIFT keypoints and compare these keypoints with keypoints of each subject with the same orientation in the keypoint database. The number of matched keypoints is obtained.
2. For each probe face, calculate the alignment error with each subject using the correlated two-phase alignment error by Eq. (5).
3. The similarity index between a probe face and gallery faces is calculated using both the number of matched keypoints and the alignment error using Eq. (6).
4. The gallery face that has the maximum similarity index in relation to the probe face is considered a matched face.

6. Results

In our experiments, the FERET database was used to evaluate the performance of the proposed algorithms. The database contains more than 14,000 faces, which are classified into several categories for different research purposes. In this study, we used the pose subset of FERET. This subset is widely used to evaluate cross-pose face recognition algorithms [11–13]. This dataset contains 200 subjects with nine poses for each subject. These poses are captured at 60°, 40°, 25°, 15°, −15°, −25°, −40° and −60°. Throughout the experiments, we used the cross-validation protocol. The database was randomly divided into two groups, Group A and Group B, which each contained 100 subjects, one group for training warps and the other for test. This process repeated 10 times and the averaged performance is used to measure the face recognition accuracy. Face images were cropped from the original FERET database to exclude hair and the background. These cropped images were then resized to a resolution of 200 × 200; the ratio of the nearest neighbor for determining the matching SIFT keypoints was set to 0.8. The specifications of the computer used to perform all experiments are as follows. All experiments were performed on an Intel Core i7-4790 with 3.6 G Hz, and 8GB of RAM running on 64-bit Windows 7 Enterprise SP1.

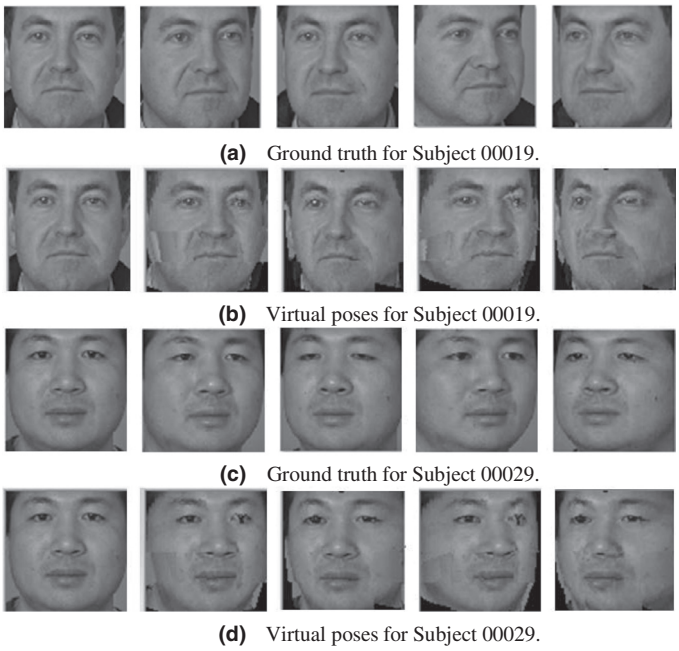

Fig. 4. Virtual views generated from a single frontal face.

Table 1
Numbers of matched keypoints and face recognition accuracy (%) for different virtual views.

	−15°/15°	−25°/25°	−40°/40°	−60°/60°
# of matching	115/123	64/72	35/32	14/22
Accuracy	99.5/99.5	96.0/97.0	67.0/65.0	20.5/19.5

6.1. Multiple virtual views

Multiple virtual views were generated from a single frontal face to enable non-frontal probe face matching with the frontal gallery face. Fig. 4 shows some of the multiple virtual views generated from a single frontal face in Group B based on the warps learned from Group A. The two subjects are presented in Fig. 4. The first and third rows are the ground truth of various poses for these two subjects from the FERET dataset; the second and fourth rows are the virtual views generated using the learned warps. Table 1 shows the quantitative comparison of virtual views and corresponding ground truth. The average number of matched keypoints is evaluated over the entire dataset between virtual views and ground truths. The accuracy illustrates the performance of face recognition using virtual views of the same pose with probe faces. Keypoints from generated virtual views behaves similarly to ones from the ground truth images. This table exhibits that our virtual views are close to the ground truth for poses within 25°, and provide discriminative feature for poses larger than 25°.

Nine virtual views were generated for each subject in the dataset. It should be noted that there are left and right orientations in the FERET dataset. As discussed in Section 3, we assumed that only the virtual views that have the same orientation as the probe face would have a positive impact on the face recognition accuracy. We compared four scenarios to verify this assumption. (1) MVV-ORIENT: a probe face was compared with virtual views that had the same orientation as the probe face. (2) MVV-SINGLE: a probe face was compared with a single virtual view that had the same pose as the probe face. (3) MVV-FULL: a probe face was compared with all generated virtual views. (4) SIFT: a probe face was compared with only a frontal face using the SIFT matching. Tables 2 shows the performance comparison of these four scenarios using the FERET database; Groups A was used for training warps and Group B was used for testing in Table 2. This table illustrates that the accuracy relationship of each scenario is as follows: MVV-ORIENT>MVV-FULL>MVV-SINGLE>SIFT.

Table 2
Performance comparison of four scenarios using the FERET database (%).

Algorithm	−15°/15°	−25°/25°	−40°/40°	−60°/60°
SIFT	99.5/99.5	96.0/95.5	68.5/59.5	19.0/18.5
MVV-SINGLE	99.5/99.5	96.0/97.0	67.0/65.0	20.5/19.5
MVV- FULL	99.5/99.5	95.5/97.0	74.5/69.0	25.0/23.5
MVV-ORIENT	99.5/100.0	97.0/98.0	79.5/79.0	26.0/26.0

This result verifies that our assumption is reasonable. Only virtual views with the same orientation as the probe face contributed to the accuracy of face recognition (MVV-ORIENT>MVV-SINGLE), whereas virtual views with other orientations not only reduced the accuracy (MVV-FULL<MVV-ORIENT) but increased the computation time. Thus, in our experiments, a probe face was compared with virtual views that had the same orientation as the probe face using SIFT. It should be additionally noted that the face recognition rate with multiple virtual views increased by approximately 10% compared to the SIFT algorithms when the pose difference was greater than 25°. The SIFT algorithm was effective when the pose difference was within 25°.

To evaluate the tolerance of the proposed method to different subsets of training face poses (corresponding to virtual views), we measure the performance of face recognition with different combinations of virtual views. As for the FERET database, it contains eight virtual views. Considering that the probe face is only compared with virtual views of the same orientation, we generated pairs of virtual views at 15°, 25°, 40°, and 60° angles for both orientations. Different combinations of these pairs of virtual views are used to test the probe face with poses ranging from −60° to 60°. In this case, the multiple virtual views do not encompass all poses in the test set. As a result, the average performance of face recognition with different subsets of virtual views is shown in Table 3. The table indicates that the efficacy of the poses is as following rank: 25°>15°>40°>60°. The virtual view of 25° is critical for both small and large angle poses. When two virtual views are used, it is recommended to include 25°, while three or more virtual views usually produce acceptable results. Virtual views from large angles, such as 60°, are generally affected by severe artifacts incurred by patch division, which degrade the performance of the SIFT algorithm.

6.2. Two-phase alignment error with correlation

The patch division process inevitably introduces the artifacts of the virtual views. Overlapping is used to reduce these artifacts. The correlation between patches was considered by using four neighboring overlapped patches and the covariance of the alignment error. Table 4 shows face recognition rates using the two-phase alignment error with correlation (AE-COR) and without correlation (AE-NCOR). This table illustrates that the correlations between patches significantly improved the face recognition accuracy. This was especially the case when the pose difference was greater than 15°; the recognition rate increased by more than 10% for pose differences of 25°, and by more than 20% for pose differences of 40°.

6.3. MVV–AE similarity

The MVV or AE independently applied are still not sufficiently accurate for cross-pose face recognition. Thus, we propose the MVV–AE framework to combine the number of matched keypoints of MVV and AE. To obtain an optimal weight, we measure the performance with various λ from 0 to 0.7 with an interval of 0.1. Table 5 shows the face recognition accuracy with various λ. A λ of 0.2 produced the optimal

Table 3
Performance comparison of different combinations of virtual views using the FERET database (%).

# of Virtual views	Combinations of virtual views (each pose represents both orientations)				
One	Comb	15°	25°	40°	60°
	Rate	70.2	72.5	66.8	65.0
Two	Comb	15°, 25	15°, 40°	15°, 60°	25°, 40°
	Rate	75.1	74.0	72.3	73.8
	Comb	25°, 60°	40°, 60°	–	–
	Rate	73.0	69.6	–	–
Three	Comb	25°, 40°, 60°	15°, 40°, 60°	15°, 25°60°	15°, 25°, 40°
	Rate	73.5	73.3	74.5	75.0
Four	Comb.	15°, 25°, 40°, 60°			
	Rate	75.6			

Table 4
Performance comparison of face recognition with correlation and without correlation using the FERET database (%).

Algorithm	−15°/15°	−25°/25°	−40°/40°	−60°/60°
AE-NCOR	95.5/94.5	71.0/77.0	25.5/28.0	15.0/9.5
AE-COR	97.0/95.0	81.5/86.0	45.5/60.5	11.0/31.5

Table 5
Average performance of face recognition with various weights, λ (%).

λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Rate	64.8	81.0	82.1	79.3	78.3	76.8	76.3	75.5

Table 6
Standard deviation of accuracies for repeated experiments using cross-validation protocol.

Algorithm	−15°/15°	−25°/25°	−40°/40°	−60°/60°	Average
MVV	1.5/1.8	3.3/2.4	4.7/6.8	3.5/5.9	3.7
AE	0.5/0.4	1.0/1.1	4.6/3.4	3.5/3.5	2.3
MVV–AE	0.5/0.4	1.0/1.1	3.6/2.9	2.6/2.6	1.8

performance among these weights. The small λ illustrates the importance of AE with regard to the accuracy even separately the MVV algorithm is more discriminative than AE as it can be seen from Table 2 and Table 4. By increasing the weight of AE, the accuracy is increased accordingly.

Fig. 5 shows the face recognition accuracy comparison of MVV, AE, and MVV–AE. This figure illustrates that MVV–AE performed better than both MVV and AE, especially when the pose difference was greater than 25°. Accuracy increased by more than 10% under these poses. The MVV and AE algorithms complement each other; accordingly, a hybrid similarity of MVV and AE achieved better results. Table 6 shows the standard deviation of accuracies based on the multiple experiments using cross-validation protocol. This result exhibits that MVV–AE not only increases the accuracy of face recognition but also gains the robustness towards different data.

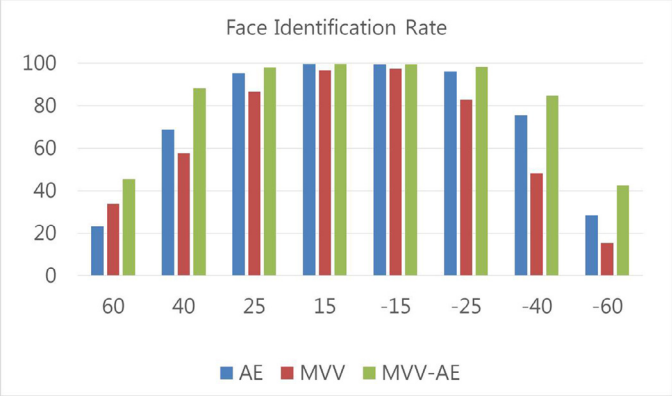

Fig. 5. Face recognition accuracy comparison of AE, MVV, and MVV–AE.

To evaluate the effectiveness of the proposed MVV–AE, we compared it with the weighted LBP [21], LGBP [3], SIFT [20], Affine-SIFT [22], stack flow [11], and OSS [16]. Local Binary Patterns (LBP) and SIFT are widely used for effective local features in many face recognition studies. The LGBP combines LBP and Gabor features, and Affine-SIFT makes SIFT features affine-invariant. A third-party dataset is required for OSS to calculate the similarity between two faces. In our experiments, OSS between the probe face and gallery face in Groups A/B was calculated by subsets in Groups B/A; each subset consisted of nine pose faces of a single subject. LDA was used to calculate the similarity between the face and a subset for the OSS algorithm.

Table 7 shows the comparison of face recognition rates and computational time with other studies. The proposed MVV–AE achieved the best performance for all poses among the algorithms to which it was compared. The LGBP demonstrated approximately 15% higher accuracy than the LBP algorithm. SIFT achieved a perfect performance for pose differences within 25°. Affine-SIFT showed only a slight improvement over SIFT. This result was due to the fact that the face is not planar; therefore, an affine transform for an entire face cannot capture the appropriate transform between poses. One-shot similarity (OSS) achieved better results than LBP for small pose changes,

Table 7
Face recognition accuracy and time comparison with other algorithms using the FERET database (%).

Algorithms	60°	40°	25°	15°	−15°	−25°	−40°	−60°	Average	Time (ms)
LBP	12.5	32.0	57.5	84.0	88.5	61.0	27.0	8.5	46.4	12
LGBP	30.0	48.5	68.0	91.5	93.5	72.5	48.5	27.0	60.0	786
OSS	27.5	47.0	72.0	89.0	91.5	69.0	40.5	21.0	57.2	752
SIFT	18.5	59.5	95.5	99.5	99.5	96.0	68.5	19.0	69.5	695
Affine-SIFT	20.5	61.5	98.5	99.0	99.5	96.0	71.0	21.5	71.0	1156
Stack flow	18.5	36.0	75.0	86.5	91.0	83.5	54.5	17.0	57.8	18
Prob. stack flow	37.0	62.0	85.0	93.0	95.5	88.0	66.5	40.0	71.0	22
AE	23.2	68.7	95.1	99.8	99.5	96.2	75.4	28.2	73.3	132
MVV	33.6	57.4	86.5	96.6	97.3	82.8	48.1	15.3	64.7	1058
MVV–AE	**45.4**	**88.1**	**98.0**	**99.8**	**99.5**	**98.2**	**84.5**	**42.5**	**82.0**	1479

such as within 25°. However, it was not capable of dealing with large pose changes. The stack flow algorithm, which is based on the Lucas–Kanade method, achieved good results for small pose changes. In general, our proposed MVV–AE achieved an approximately 20% face recognition accuracy increase over LBP-based algorithms (LBP, LGBP, OSS) and a 10% increase over SIFT-based algorithms (SIFT, Affine-SIFT) and Lucas–Kanade algorithms (stack flow). In particular, our proposed MVV–AE algorithm achieved 95% face recognition accuracy when pose differences were less than 40°. The computational time shows the time for testing with a gallery face. The proposed MVV–AE exhibited the worst time, while it is worthy and acceptable due to the higher accuracy.

7. Conclusion

In this paper, a novel face recognition framework based on multiple virtual views and alignment error was developed and implemented. The method enables a non-frontal probe face to be compared with a frontal face in a gallery. To overcome the problem of differing poses, a frontal face is transformed into multiple virtual views using learned warps between poses. SIFT is used to calculate the number of matched keypoints between the probe face and virtual views. Furthermore, a two-phase alignment error was proposed to capture the pose and individual alignment error between faces, while correlations between patches are used to improve accuracy. Finally, the number of matched keypoints and the alignment error between the probe face and gallery face are combined to calculate the similarity between them. Experimental results showed that multiple virtual views significantly increased the cross-pose face recognition rate compared to a single SIFT algorithm. Furthermore, combining the alignment error with correlation, our proposed method achieved impressive results when it was compared to other algorithms.

Acknowledgments

This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST; no. 2012R1A2A2A03).

References

[1] NIST. Video Challenge Problem Multiple Biometric Grand Challenge Preliminary Results of Version 2, 2009. 1.
[2] G. Hua, M.H. Yang, E.L. Miller, Y. Ma, M. Turk, D.J. Kriegman, T.S. Huang, Introduction to the special section on real-world face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1921–1924.
[3] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2005, pp. 786–791.
[4] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification, in: Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3025–3032.
[5] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Yi Ma, Robust face recognition via sparse representation,, IEEE Trans. Pattern Anal. Mach. Intell 31 (2009) 210–227.
[6] D. Chen, X. Cao, L. Wang, F. Wen, J. Sun, Bayesian face revisited: a joint formulation, in: Proceedings of European Conference on Computer Vision (ECCV), 2012, pp. 566–579.
[7] X. Cao, D. Wipf, F. Wen, G. Duan, J. Sun, A practical transfer learning algorithm for face verification, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2013, pp. 3208–3215.
[8] Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf, Deep Face: Closing the gap to human-level performance in face verification, in: Proceedings of International Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[9] H.T. Ho, R. Chellappa, Pose-invariant face recognition using Markov random fields, IEEE Trans. Image Process. 22 (2013) 1573–1584.
[10] S.R. Arashloo, J. Kittler, Energy normalization for pose- invariant face recognition based on MRF model image matching, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1274–1280.
[11] A.B. Ashraf, S. Lucey, T. Chen, Learning patch correspondences for improved viewpoints invariant face recognition, in: Proceedings of International Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[12] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1–12.
[13] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, S. Shan, Morphable displacement field based image matching for face recognition across pose, in: Proceedings of European Conference on Computer Vision (ECCV), 2012, pp. 102–115.
[14] S.J.D. Prince, J.H. Elder, J. Warrell, Fatima, Tied factor analysis for face recognition across large pose differences, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2008) 970–982.
[15] A. Sharma, M.A. Haj, J. Choi, L.S. Davis, D.W. Jacobs, Robust pose invariant face recognition using coupled latent space discriminant analysis, Comput. Vis. Image Underst. (2012) 1095–1110.
[16] L. Wolf, T. hassner, Y. Taigman, Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1978–1990.
[17] A. Li, S. Shan, W. Gao, Coupled bias–variance tradeoff for cross-pose face recognition, IEEE Trans. Image Process. 21 (2012) 305–315.
[18] Y. Zhang, M. Shao, E.K. Wong, Y. Fu., Random faces guided sparse many-to-one encoder for pose-invariant face recognition, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 2416–2423.
[19] M. Kan, S. Shan, H. Chang, X. Chen, Stacked progressive auto-encoders (SPAE) for face recognition across poses, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 4321–4328.
[20] G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110.
[21] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 2037–2041.
[22] J.M. Morel, G. Yu, ASIFT, A new framework for fully affine invariant image comparison, SIAM J. Imaging Sci. 2 (2009) 438–469.
[23] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, Image Vis. Comput 16 (1998) 295–306.